Anal Chem 48:1368-1375

*** REVISED MANUSCRIPT - NEW PAGINATION ***

EXTRACTION OF MASS SPECTRA FREE OF BACKGROUND AND NEIGHBORING COMPONENT CONTRIBUTIONS FROM GAS CHROMATOGRAPHY/MASS SPECTROMETRY DATA

R. G. Dromey*, Mark J. Stefik, Thomas C. Rindfleisch, and Alan M. Duffield

Departments of Computer Science, Genetics, and Chemistry, Stanford University, Stanford, Calif. 94305.

* Current Address
Research School of Chemistry
Australian National University
Canberra, A.C.T.
Australia

BRIEF:

A computer-based method has been developed for extracting from GC/MS data, component mass spectra that have had background removed and which have been corrected for contributions from neighboring components.

ABSTRACT:

A computer-based method has been developed for extracting from GC/MS data, component mass spectra that have had background removed and which have been corrected for contributions from neighboring components. Components are detected in the data by a pair of histograms which characterize the positions of mass fragmentogram peak modes. Tabular least squares peak modeling is then employed to facilitate the resolution of multiplet fragmentogram complexes and to correct the data for background contributions and for peak saturation. These corrections are all essential prerequisites for obtaining representative mass spectra from GC/MS data of complex mixtures on a routine basis. Using the present approach, components that elute within two scan times can be detected and resolved into their respective contributions.

INTRODUCTION:

With the increasing application of gas chromatography/mass spectrometry (GC/MS) systems to mixture component identification in biomedical research (1, 2) and other areas (3), it has become important to be able to systematically isolate and identify minor components in the complex mixtures being analyzed. Because of instrumentation limitations, the mass spectra obtained from a GC/MS analysis of a complex mixture are often markedly different from the spectra of the corresponding pure compounds. Differences may be caused by contributions from unresolved neighboring components during partial separation and also from GC septum and column bleed. These extraneous contributions may severely distort the relative intensities of ions in the mass spectrum of a particular component as well as contribute peaks that are not characteristic of the component being examined. Characterization and removal of these spurious ion contributions is especially important in the analysis of minor constituents where the mass spectra of interest may be substantially masked or distorted.

Our objective has been to implement a solution to these problems which is general and can systematically and reliably resolve GC/MS data with a minimum of human intervention. At the same time we have constrained the design so that the programs can run on a laboratory mini-computer. The first of these objectives has necessitated the use of a relatively complex mathematical treatment of the GC peak profile analyses as compared to that previously reported by Biller and Biemann (4). Both the present approach and that of reference (4) are based on analyses of mass fragmentogram profiles (4, 5, and 6), a method which has been in use in various laboratories for a number of years, including our own. The method described here, however, differs substantially in the extraction of information from the profiles and thereby avoids several serious limitations inherent in the

system described previously (4). By using tabular models of the elutant peak shapes derived from the raw data combined with an approximation to the GC background, and by deploying the elutant location and multiplicity information gained in analyzing individual fragmentogram profiles to assist in analyzing the others, we can achieve significant advantages in the quality of the reduced data as detailed in the following sections. These include better final GC resolution, the proper assignment of ions to resolved elutant spectra (whether or not they are shared between neighboring components), more accurate spectral amplitudes free from background contributions, and the recovery of usable information from damaged data as in saturated peaks. We feel these improvements are critical to a system which can reliably extract component spectra of sufficiently high quality from GC/MS runs to enable more definitive library matching, easier human interpretation of unknowns, and even the addition of extracted spectra to the library as authentic spectra. In our experience these are essential assets for a GC/MS data system which is to be routinely applied in medical research and amply justify the complexity of the analysis.

EXPERIMENTAL:

The GC/MS computer system used in this investigation consists of a Finnigan 1015 Quadrupole mass spectrometer interfaced to a PDP11/20 mini-computer system for data acquisition. In one frequent mode of operation a complete mass scan (from mass 40 to 450) is completed each 3.7 seconds and 600 consecutive mass spectra are collected during a typical GC/MS analysis. Our initial experience of comparing the experimental mass spectra from a complex GC/MS analysis with a library of known mass spectra produced very poor results due to contamination of the experimental data by spectra of column bleed and of neighboring, unresolved

components. Tolerable matches were only achieved when a component was present in large quantity in the GC/MS analysis. In order to overcome these problems, we have developed a computer program capable of systematically extracting from the raw GC/MS data, spectra representative of the pure elutant compounds.

The raw mass spectrum (Figure 1a) of indole acetic acid 3-methyl ester obtained from a GC/MS analysis of the acidic fraction (after methylation) of human urine typifies this situation. This component elutes at or near spectrum number 492 in the total ion plot (TIC) shown in Figure 2. Closer examination of Figure 2 shows that this component is submerged both in mass spectral contributions from neighboring components and background due to GC column bleed. For comparison a library spectrum (7) of indole acetic acid 3-methyl ester is shown in Figure 3. Figure 1b shows how after processing the raw GC/MS data by the method described below we can retrieve a high quality mass spectrum of indole acetic acid 3-methyl ester free from the environmental perturbations present in Figure 1a.

In the systematic analysis of GC/MS data the problem is then first to detect where in the GC trace each component shows its maximum ion intensity and then to extract from these regions representative spectra of each of the detected components. The extracted mass spectra should be as free as possible from intensity distortions, relative to their library counterparts, and from the presence of ions that are not present in the reference (library) mass spectrum of each component (e.g., peaks from either neighboring components or gas chromatographic column bleed).

DESCRIPTION OF METHOD:

To obtain a reliable solution to these problems, it is necessary to analyze

a number of spectra on either side of the ion current maximum for each elutant. A basic assumption of our approach is that the mass spectra of two neighboring unresolved elutants can be distinguished; that is, there exist some masses for which ions occur in the mass spectrum of one component but not in the other and vice versa. A schematic representation for two closely spaced elutants is given in Figure 4. By locating the "resolved" or singlet fragmentogram peaks at such masses (detected on the basis of profile morphology) one can infer directly the positions of the elutants present and derive tabular models of the individual peak shapes which can be used to separate the unresolved fragmentogram complexes. The use of tabular peak models derived from the data itself accurately accommodates the a priori unknown peak profiles of particular elutants without solving for multiparameter, non-linear model functions. Since the data are sampled often enough to satisfy the sampling theorem (8), these tabular models contain the necessary information to reconstruct the continuous peak envelope and can therefore be used as if they were continuous analytical models. For the typical peak shapes encountered, the collection of 5-10 mass spectra per singlet elutant peak represents a sampling frequency greater than twice the Fourier bandwidth of the peak. In addition, the mass by mass analysis of the fragmentogram peak complexes facilitates the mass dependent subtraction of background. (The large variation in background levels for different masses is a function of both the type of GC column used and the mixture being analyzed).

By addressing the problem in this way we have been able to produce accurate intensity information for the processed mass spectra and simultaneously distinguish with greater confidence which masses contribute to particular elutant spectra. We have been able to distinguish reliably elutants coming off within one and a half to two spectral scan times of each other. The succeeding sections

discuss in more detail the procedures used to detect and resolve the mass spectra of unique elutants.

Detection of elutants in GC/MS data:

Elutant detection involves finding the location of each mixture component in the GC/MS data, even if it does not have a corresponding peak maximum in the overall total ion current trace. Ideally for a given elutant, the fragmentograms for all its ion masses will show maxima at the same time and in practice this holds for well-resolved materials. However, for partially resolved mixtures, the complicating factors of peak overlap and background contributions can cause fragmentogram maxima for neighboring components to show significant variation in their positions on the time axis. Reliable position information for each elutant is best derived from the fragmentogram profiles containing singlet peaks for that elutant, that is, from fragmentograms at those spectral masses unique to the elutant relative to its neighbors. Thus elutant detection may be viewed as locating such peaks.

The basic approach used for elutant detection is to compute two histograms of candidate singlet peak positions and to select as elutant locations significant histogram maxima. The first histogram measures the number of singlet mass fragmentogram profiles which reach maxima in each time interval. The second histogram measures the total singlet ion intensity above background at these maxima. These two types of histogram contribute complementary information for judging elutant locations. At a given elution time, the histograms include fragmentogram peak maxima from all masses over seven spectra. The position of each maximum is determined by a parabolic least squares interpolation about the top five points in the sampled peak data. If the intensities of the five points

contributing to the maximum are Y_{-2}, Y_{-1}, Y_0, Y_1, and Y_2, then the expression for the time coordinate of the maximum is

$$t = \frac{7\,(2\,Y_{-2} + Y_{-1} - Y_1 - 2\,Y_2)}{10\,(2\,Y_{-2} - Y_{-1} - 2\,Y_0 - Y_1 + 2\,Y_2)}$$

The time coordinates of maxima are estimated to one third of the time to collect each spectrum in order to separate very close neighbors. Because we measure peak locations to one third of a spectral scan time, appropriate shifts are also included to account for the fact that higher masses are measured later in each spectral scan than lower masses. To build these histograms, the program examines the profiles of each mass fragmentogram in the data. Only peaks with intensities above a prescribed threshold are added into their appropriate time positions in the histograms. Peaks that are obvious multiplets (multiple extrema) are not incorporated into the histograms but are marked for later resolution. After all of the histogram information is collected for a given region, compounds are defined to be detected at locations where both the intensity and peak count histograms show maxima that are above a threshold. This statistical approach, looking for "clusters" of fragmentogram peaks in the histograms, does not depend upon a correct decision for each peak but rather on a preponderance of good decisions looking over all of the data. It will fail to resolve elutants very close together which do not have enough distinguishing mass spectral components as described above. In general, however, using this approach we are able to detect and resolve spectra reliably that elute with a separation in time as small as one and a half to two spectral scan times. (Two scan times corresponds to 25 percent of a typical GC peak width at the scan rate we use). Elutants this close often do not show multiple extrema in the fragmentogram profiles of masses common

to both spectra and could not be separated properly except for this type of procedure. If two elutants are separated by less than 1.5 scan times, resolution becomes more uncertain depending on their relative concentrations and mass spectral distinctness.

Estimation of spectral intensities and background for well resolved elutants:

Once the locations of elutants in the GC effluent have been determined, we proceed to compute a resolved spectrum for each material. To illustrate the principles involved in spectral amplitude and background estimation, we consider the simple case of an elutant that is well separated in time from its nearest neighbors. This analysis will be extended to the more complicated case of multiplet resolution in a later section. By "well-separated" we imply only that there are no maxima in the elutant detection histograms for three or four spectra on either side of the elutant under consideration. In such a situation, each of the mass fragmentogram profiles in the vicinity of the elutant will consist of a background on which is superimposed a peak with amplitude representative of the elutant spectral component at that mass. The background (due to both GC column bleed and possible tailing from nearby, high-concentration elutants) is distinguished from the elutant peak by the fact that it varies much more slowly with time. Reasonable estimates can be made by assuming that for any particular mass fragmentogram the contribution is linear (see Figure 5). This approach to background determination, using the actual fragmentogram characteristics in the vicinity of each elutant, automatically tracks changes in the bleed levels observed during a run. It should be noted that our model is a first order approximation subject to some error. A more accurate approximation would involve representing the background variations over a larger span of spectral scans than

we are able to manage with the current program organization and computer memory limitations. We feel that the linear estimate is justified, however, in that it produces results within the error limits due to other data uncertainties.

To complete the estimation process we use a model peak to determine the contribution of each mass fragmentogram to the elutant spectrum. Much work has been done on the analytic approximation of gas chromatographic peak shapes (9, 10). Our experience has been that relatively simple models do not adequately approximate the range of shapes encountered and more complex models require large amounts of computing to determine model parameters. Noting that a separate model must be developed for each elutant and with a view toward obtaining the peak shape and definition necessary for multiplet resolution within reasonable computing resources, we have approached the problem by using tabular peak models taken from the data itself. Such models, defined at discrete sample points, can be evaluated at any required intermediate point by interpolation (since the sampling theorem is satisfied) and automatically reflect any peak asymmetries which may be present. For a given elutant, the model will be independent of mass, assuming that relative molecular fragmentation probabilities do not change with elutant pressure within the mass spectrometer. A number of criteria should be satisfied by the tabulated model peaks. They should be singlet peaks superimposed on as small a background as possible and they should be relatively intense in order to ensure a good signal-to-noise ratio and good definition of peak skirts.

Candidate singlet peaks may be distinguished from doublet or background peaks by the feature that they are relatively sharp. One way to measure peak 'sharpness' is to use a logarithmic rate function defined as follows:

$$\text{rate} = \sum_{t=1}^{3} \left[\frac{(Y_{t-1} - Y_t)}{Y_t} + \frac{(Y_{-(t-1)} - Y_{-t})}{Y_{-t}} \right]$$

where the Y_t are evaluated at equal scan widths at each side of the mode of the peak. It can be seen that this rate will be large for peaks which are sharp and smaller for peaks which are broad. The rate as defined is also independent of amplitude for peaks of identical shape. A peak with a computed rate below a threshold appropriate to the experimental conditions is considered to be either an artifact of the gas chromatograph (background peak) or a multiplet and is not included in the detection histograms.

During the process of computing the detection histograms, a list is kept of the unimodal fragmentogram peaks having the highest rate factors in the region under analysis. When a component is detected in a given region, a model peak is then immediately in hand that can be used in the peak height estimation and background removal process. The local minima just on either side of the model peak are used as estimates of the local background (a straight line through the greatest of these minima is removed before the model peak is used for analysis). The selection of the peak with the highest rate factor as our model peak has worked well in producing models which are singlets and suffer least from interference by background and neighboring fragmentogram peaks.

Given the fragmentogram peak model for this case of a well-separated elutant, we can now correct the individual mass fragmentograms for background and estimate true mass spectral intensities for the elutant. For the fragmentograms exhibiting peak maxima "near" the location of this elutant (see below for detailed selection criteria), each peak in the set is quadratically interpolated to align it on a common time origin (this removes the time shift between collection of low and high mass data). This is done by fitting a parabola through successive groups of three points near the peak mode and interpolating to give four equally spaced points about the mode, separated by one spectral scan

time. With the peaks in this standard form they are ready for the least squares analysis below. Assuming a linear background model over the region of 5 to 10 scan intervals under consideration, the local background B_t at time t is represented by

$$B_t = c + d\,t$$

where c is the background offset and d is its slope. The interpolated elutant peak model is normalized to unit area and has amplitudes P_t at times t. Then for a given mass fragmentogram, the amplitude of the actual fragmentogram profile Y_t at time t can be represented by

$$Y_t = p\,P_t + (c + d\,t)$$

where p measures the elutant amplitude above background. Note that this model assumes a superpositon principle based on the earlier mention of mass spectrometer fragmentation probabilities and a linear encoding of ion current information. If ion current data are obtained from non-linear electronic systems or read from film, the peak model itself would be amplitude dependent. From this model we can derive a least squares estimate for the elutant amplitude p and the background parameters c and d by minimizing the error function

$$E = \sum (Y_t - p\,P_t - c - d\,t)^2$$

The error is minimum when

$$\frac{\partial E}{\partial p} = \frac{\partial E}{\partial c} = \frac{\partial E}{\partial d} = 0$$

which yields three linear equations in three variables

$$\begin{aligned}
p \sum P_t^2 + c \sum P_t + d \sum t\,P_t &= \sum P_t\,Y_t \\
p \sum P_t + c\,N + d \sum t &= \sum Y_t \\
p \sum t\,P_t + c \sum t + d \sum t^2 &= \sum t\,Y_t
\end{aligned}$$

where N is the number of points in the summation. It is worth noting that this method, using a tabular model peak derived from the data and elutant locations obtained from the detection histogram analysis, reduces the calculation for each mass spectrum intensity to the solution of a set of linear equations. Specifically this avoids iterative methods for determining the parameters of a theoretical peak model and for determining elutant time positions. From the solution of these equations for the value of p, we get the spectral intensity for each mass. This analysis is applied to all mass fragmentograms with maxima near the elutant location to obtain the intensity-corrected spectrum.

Fragmentograms are selected for this analysis on the basis of several criteria. Given the nominal elutant position from the detection histogram analysis, a fragmentogram is excluded (mass spectrum assigned zero intensity) if it has no local peak maximum or if its maximum is displaced from the reference elutant position by more than two thirds of a spectral scan time on either side. Each fragmentogram peak meeting this test must also have an acceptably high rate factor, to be included in the analysis. For peaks of masses greater than 200 amu, we require a rate factor greater than 25 percent of the rate for the model peak. This restriction is useful for eliminating contributions caused by peaking in column bleed components. In carefully examining GC/MS data sets, we can observe that masses characteristic of the spectra of column bleed components show maxima in their mass fragmentograms just prior (one to two spectra) to the elution of an actual component. In essence the component appears to "push" the bleed out ahead of itself. Because these peaks are formed by a different process than normal elutant mass fragmentogram peaks, they usually have a much broader shape. Consequently their rate factors will be significantly reduced and they can be eliminated by the rate threshold criterion. The combination of the

fragmentogram peak location criterion together with the minimum rate criterion effectively discriminates against extraneous contributions to the intensity-corrected spectra without removing authentic mass peaks.

Extraction of poorly separated elutant spectra:

Many instances arise in the analysis of GC/MS data where two or more elutants are poorly resolved by the gas chromatograph. The resulting mass spectra in such a region exhibit ion intensity distortions which reflect the interactions (overlap) between adjacent elutants in addition to the ion contributions of background. The extension of the above procedures to the general case is not difficult. Through the histogram detection and model procedures, one can extract normalized peak models P, Q, R, ... for the various elutants present. Then with the assumption of a linear background, the fragmentogram profile Y can be approximated by minimizing the error function

$$E = \sum (Y_t - p\,P_t - q\,Q_t - r\,R_t - \ldots - c - d\,t)^2$$

with respect to the elutant amplitudes p, q, r, ..., and the background coefficients. Sets of linear equations result for each mass to extract the resolved spectra. In practice, we have not implemented this full procedure beyond the doublet case. Through the following approximations, reasonable results are achievable within available mini-computer resources. Using the histogram method described earlier, neighboring elutants are handled with a 'lookahead' procedure. That is, information about an elutant that has just been detected is stored and the detection algorithm is applied to the data in the immediate neighborhood by extending the range over which the detection histograms are calculated. If by including this extended region an additional elutant is

detected, we record the position of its mode, select a model peak for this second elutant using the rate criterion, and initiate a doublet resolver algorithm. At present, the extended histograms project four spectral scan widths beyond the position where the first elutant of the multiplet was detected. The same criteria are applied as in the singlet case to decide which fragmentogram peaks belong to the pair of detected components. The model used to process the composite fragmentogram peaks (many of which may be singlets belonging to either elutant) assumes that there are two overlapping peaks superimposed on a linear background. A schematic representation of this model is given in Figure 6. The doublet model represents an oversimplification of some situations as, for example, in the case where 3 components elute within a very brief interval. It provides, for the most part, however, acceptable accuracy for resolving multiplet GC peaks and estimating individual mass spectra for the respective components. As indicated above, a least squares fit of the two peak models P and Q to the fragmentogram profile Y with a linear background may be described by the equation

$$Y_t = p\,P_t + q\,Q_t + c + d\,t$$

At minimum error this yields four linear equations as follows:

$$\begin{aligned}
p\sum P_t^2 &+ q\sum P_t Q_t + c\sum P_t + d\sum t\,P_t = \sum P_t Y_t \\
p\sum P_t Q_t &+ q\sum Q_t^2 + c\sum Q_t + d\sum t\,Q_t = \sum Q_t Y_t \\
p\sum P_t &+ q\sum Q_t + c\,N + d\sum t = \sum Y_t \\
p\sum t\,P_t &+ q\sum t\,Q_t + c\sum t + d\sum t^2 = \sum t\,Y_t
\end{aligned}$$

In cases where peaks are actually singlet peaks they should yield zero for the amplitude of the missing component. In practice for such cases the amplitude of the second component is a very small positive or negative value which is representative of how well the model fits the data. Amplitude results for masses

that belong to the second component of the doublet are stored temporarily until this component is moved into the processing window at which time they are incorporated into the analysis of the newly detected component.

Reconstruction of saturated peaks in elutant spectra:

From a practical viewpoint, a fairly common occurrence in GC/MS data collection systems is the problem of mass peak saturation. Saturated peaks occur when the concentration of a component in the ion source is such that for one or more ion masses the detection system analog-to-digital converter becomes overloaded. Saturated peaks are easily detected because of their characteristic flat tops which have an amplitude determined by the overload limit of the detection system (e.g., the saturation value is 4095 if a twelve-bit analog-to-digital converter forms part of the detection system). Figure 7 represents a saturated mass fragmentogram peak superimposed on a linear background.

If we are to obtain accurate intensity profiles for component spectra that include one or more saturated mass peaks it is essential that we reconstruct these saturated peaks in a way that will give a relatively precise estimate of their true amplitudes. A convenient way to do this in the singlet case is to use the least squares model that we derived in the preceding sections. To actually apply it for reconstruction of saturated mass fragmentogram peaks we need to make a small modification to the equations. Instead of summing over all the points in the peak, we sum over only those points that are not saturated in the fragmentogram, and let N represent the number of such points. As an estimate of the peak mode we use the mode of the intensity histogram for the component being analyzed. An example of reconstruction of a mass spectrum with saturated ion intensities is given in Figure 8. Figure 8a shows a saturated spectrum of

tetracosane and Figure 8b is the corresponding reconstructed spectrum. It is clear that the reconstructed spectrum will give a far better match with a library spectrum than the saturated spectrum which is badly saturated at masses 43, 57, and 71.

Before leaving the discussion of saturation we should point out that we have not in practice extended the procedure for saturation correction of singlet peaks to the doublet case as we believe that it would be inadequate for reliable intensity estimates. If too many points are overloaded there will be insufficient data to accurately estimate the amplitude of each multiplet component. Despite such correction algorithms, there is no substitute for the collection of good quality raw data at the start.

RESULTS AND DISCUSSION:

The program based on the algorithm outlined in the preceding sections has been tested on a wide variety of biological samples. It fits comfortably into a DEC PDP 11/45 computer (with 28k of memory) and takes approximately 8 minutes to analyze a raw GC/MS data set of 600 mass spectra (scanned from masses 40 to 450). Much of this time is spent in reading the raw data from the disk and other input-output operations. Copies of the program, which is written in FORTRAN, are available from the authors. Currently this program forms part of an automated analysis system for the GC/MS analysis of urine and blood samples. The program reduces the raw GC/MS data set of approximately 600 spectra to a set of about 60 resolved elutant spectra which are then matched against a library of mass spectra of biological compounds. This whole process takes about 20 minutes and produces an analysis of the sample, with known compounds in the mixture identified and the remaining unknown set marked for further study by chemists or other DENDRAL programs (11, 12, and 13).

In evaluating performance of the program, a major issue is how well it is able to detect elutants in the data. The vertical bars on the TIC (Figure 2) indicate all the places where the program detected and isolated a component from the raw GC/MS data. The program's power of detection is illustrated by the elutants detected near spectrum numbers 492 and 543 in the total ion current plot shown in Figure 2. Although there is no evidence of maxima in the TIC in the regions near 492 and 543, the program was able to detect and isolate quality spectra of known compounds. Indole acetic acid methyl ester (Figure 1) is found near spectrum number 492 and is clearly submerged in background and overlapping contributions, while 4-methoxybenzoyl glycine methyl ester (Figure 9) detected near spectrum 543 in Figure 2 is almost completely masked by the large concentration of hydrocarbon internal standard (n-C24H50) which maximizes at spectrum number 545. The detectability of unresolved elutants is clearly a function of their amplitude relative to neighboring components and background. One way to characterize this is to measure the ratio of the total ion intensity (sum of the mass spectrum amplitudes) in the resolved spectrum compared to that in the unprocessed spectrum including background and overlap effects. The mass spectrum of the processed component at spectrum number 492 comprises only 4 percent of the total raw ion current. It can be expected that there will be problems detecting components with an ion current ratio that falls much below a level of 4 percent. Also if two compounds elute within less than 1.5 to 2 spectral scan times of one another, there is an increasing chance that the program will make the wrong decision as to whether there is one or actually two elutants present. Such errors are also dependent on the ion current ratio between adjacent elutants and the stability with which peak positions can be determined.

As an example of doublet resolution consider the region near spectrum numbers 317 and 318 in Figure 2. The program detects that there are two elutants and Figures 10 and 11 illustrate the raw and resolved spectra at these locations. The spectrum in Figure 10b is a good representation for 4-methoxyphenylacetic acid methyl ester. The other component is an unknown aromatic ester. The spectrum at index number 308 (Figure 2) is shown in Figure 12 and is another example of how the program detects and corrects a small isolated component (furan dicarboxylic acid dimethyl ester).

We have evaluated the efficiency of background removal for singlet elutants by examining their mass fragmentograms. After calculating the least squares peak and background levels, we concluded that the computed results are consistent (5 - 10 percent) with human estimates. They tended to be less accurate for very weak peaks whose shapes were more sensitive to noise distortions.

For the multiplet case, where the peak profiles can be considerably more complex, there is a stronger possibility that the model will not produce accurate amplitude information. In such cases, as when there are three rather than two elutants present, there is a danger that background contributions will be incorrectly estimated. We feel however that use of a more complex model for triplets is not likely to be able to guarantee much greater precision because of the limited accuracy of the raw data itself. Problems most frequently occur when a small amount of an elutant occurs just prior to, or just after, an elutant of high concentration. The intensities of peaks in the small elutant that are common to the large elutant tend to be less accurately calculated than singlet peaks and sometimes may even be discarded as negligible if their intensity relative to the large peak falls much below 10 percent.

Comparison with library mass spectra has indicated that correction of

intensities for saturated singlet peaks is satisfactory. However, as expected the accuracy of the calculation decreases as peaks become more heavily saturated. In our case we are working with model peaks that extend over nine points (i.e., nine scan widths). If more than four of a peak's nine points are saturated we can expect that its estimated intensity will have only limited accuracy because there is insufficient information left to accurately characterize its shape.

An example will illustrate the overall performance of the program. Let us consider in detail the spectrum of the elutant detected near spectrum number 492 (Figure 2). A comparison of the resolved spectrum Figure 1b with a library spectrum (Figure 3) shows that the basic spectral intensity profiles are very similar even including the very low intensity ion of mass 89. Some very small ions (of intensity less than 5 percent relative abundance) are absent from the resolved spectrum because they have been lost in the background noise. It is worth noting that there are no peaks present in the resolved spectrum that are not in the library spectrum, that is, the extraneous mass spectral peaks in the raw data including peaks at masses 105, 253,and 315 are not included in the resolved spectrum. The relative intensities of the mass spectral peaks at masses 51,62, 65, and 77 have been changed significantly from their levels in the raw data. This illustrates the importance of correcting the intensities for background. The mass spectral peaks at masses 51, 52, 63, 78, and 129 appear to maximize near spectrum 496 in the raw data instead of spectrum 492 because of the overlapping contributions of a poorly resolved elutant.

Conditions arise in the raw GC/MS data for which it is not possible to extract unambiguously high quality mass spectra. One case is when the elutant-to-background ratio falls significantly below 5 percent. In these cases, the very weak intensity ions, including isotope ions, usually do not appear in the

resolved mass spectra. The other difficulties arise when it is not possible to detect the presence of multiple elutants because they occur within less than one mass spectrum scan time of each other. In this case, the processed spectrum represents the mixture of the two elutants.

In general, we have found that the present system is capable of detecting and isolating high quality representative mass spectra in GC/MS experiments involving complex biological mixtures.

ACKNOWLEDGEMENTS:

We wish to thank Dr. B. E. Blaisdell of Juniata College, Huntingdon, Pennsylvania for his support on an exploratory alternative approach to the present problem.

CREDITS:

This work was supported by grants (Nos. RR-612 and GM-20832) from the National Institutes of Health and (No. NGR-05-020-632) from the National Aeronautics and Space Administration.

REFERENCES:

1. R. A. Hites and K. Biemann, Anal. Chem., 42, 855 (1970).
2. C. C. Sweeley, N. D. Young, J. F. Holland, and S. C. Gates, J. Chrom. 99, 507 (1974).
3. W. H. McFadden, Techniques of Combined Gas Chromatography/ Mass Spectrometry: Applications in Organic Analysis, Wiley Interscience, London (1973).
4. J. E. Biller, and K. Biemann, Anal. Letters., 7, 515 (1974).
5. R. E. Summons, W. E. Pereira, W. E. Reynolds, T. C. Rindfleisch, and A. M. Duffield, Anal. Chem., 46, 582 (1974).
6. R. N. Stillwell, 22nd Annual ASMS Conference - Mass spectrometry, Philadelphia, 1974 p. 454.
7. Mass Spectra of Compounds of Biological Interest, U. S. Atomic Energy Commission Report No. TID - 26553, S. P. Markey, W. G. Urban, and S. P. Levine (Eds).
8. See for example: R. S. Ledley, Digital Computer and Control Engineering, p. 742, McGraw-Hill Book Co., Inc. (1960).
9. E. Grushka, M. N. Myers, and J. C. Giddings, Anal. Chem., 42, 21 (1970).
10. C. D. Scott, D. C. Chilcote, and W. W. Pitt, Clin. Chem., 16, 637 (1970).
11. A. Buchs, A. B. Delfino, A. M. Duffield, C. Djerassi, B. G. Buchanan, E. A. Feigenbaum, and J. Lederberg, Helv. Chim. Acta, 53, 1394 (1970).
12. R. E. Carhart, D. H. Smith, H. Brown, and C. Djerassi (in press in J. Amer. Chem. Soc.).
13. R. G. Dromey, B. G. Buchanan, D. H. Smith, J. Lederberg, and C. Djerassi, J. Org. Chem. 40, 770 (1975).

FIGURE CAPTIONS:

1. (a) Spectrum of indole acetic acid 3-methyl ester from a GC/MS analysis of human urine before processing.

   (b) Resolved spectrum of indole acetic acid 3-methyl ester.

2. Total ion current plot for a GC/MS analysis of a urine sample. Components were found at vertical bar marks on TIC.

3. Mass spectrum of indole acetic acid 3-methyl ester taken from a library of biological compounds.

4. A schematic representation of the set of partial mass fragmentograms for two closely spaced elutants. Components A and B have some masses in common.

5. Model for a singlet peak located on a linear (sloping) background.

6. Model for a mass fragmentogram containing a doublet superimposed on linear sloping background.

7. Model for a saturated singlet peak on a linear background.

8. (a) Saturated unprocessed mass spectrum for tetracosane.

   (b) Spectrum of tetracosane after processing and correction for saturation.

9. (a) Mass spectrum of 4-methoxybenzoyl glycine methyl ester before processing.

   (b) Resolved spectrum of 4-methoxybenzoyl glycine methyl ester.

10. (a) Mass spectrum of 4-methoxyphenylacetic acid methl ester before processing.

    (b) Resolved mass spectrum of 4-methoxyphenylacetic acid methyl ester.

11. (a) Spectrum of an unknown aromatic ester before processing.

    (b) Resolved spectrum of unknown ester in Figure 11a.

12. (a) Spectrum of furan dicarboxylic acid dimethyl ester before processing.

    (b) Resolved spectrum of furan dicarboxylic acid dimethyl ester.

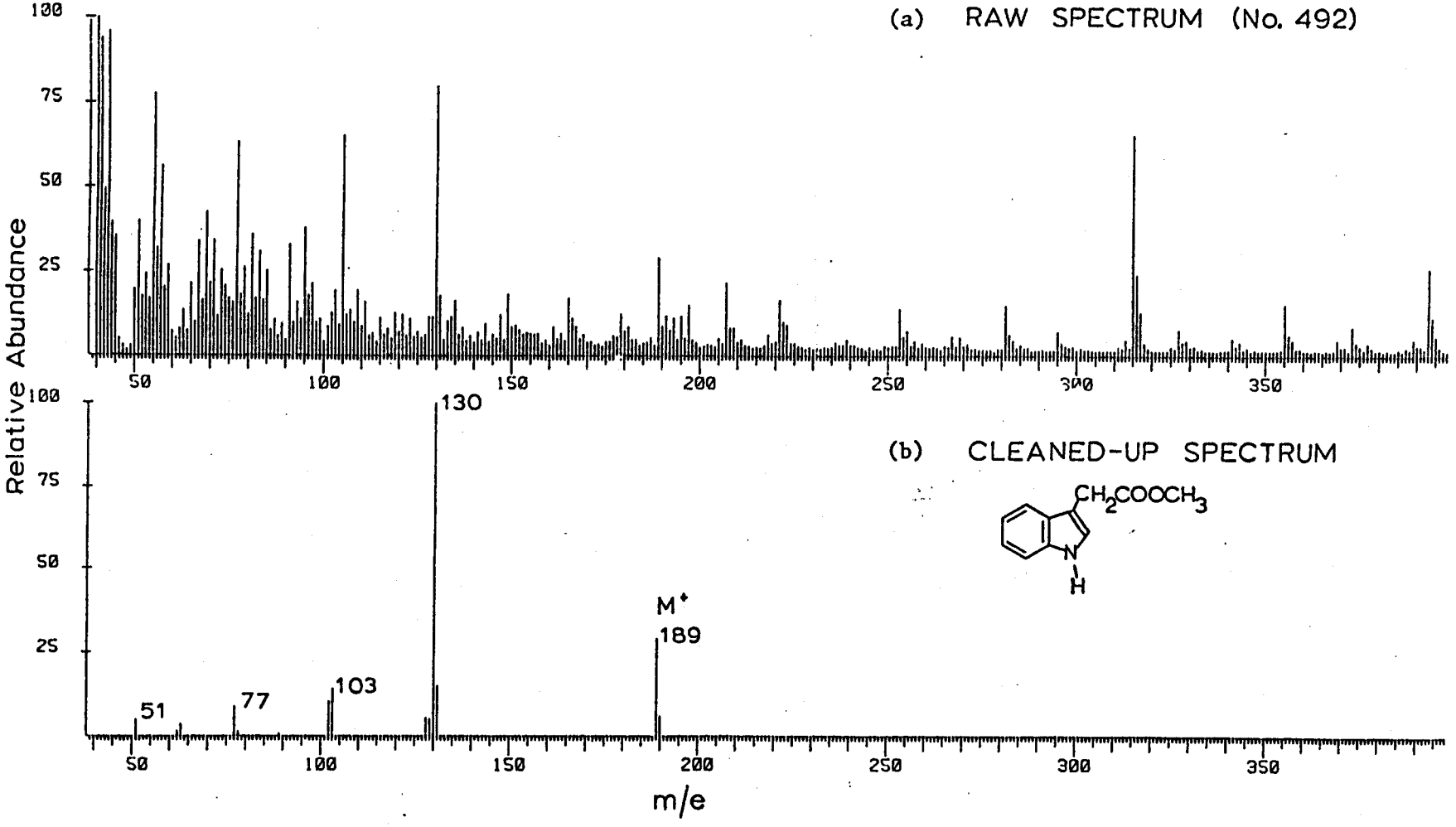
(a) RAW SPECTRUM (No. 492)
(b) CLEANED-UP SPECTRUM
Relative Abundance
m/e
130
M+
189
103
77
51
CH2COOCH3
N
H

FIGURE 1

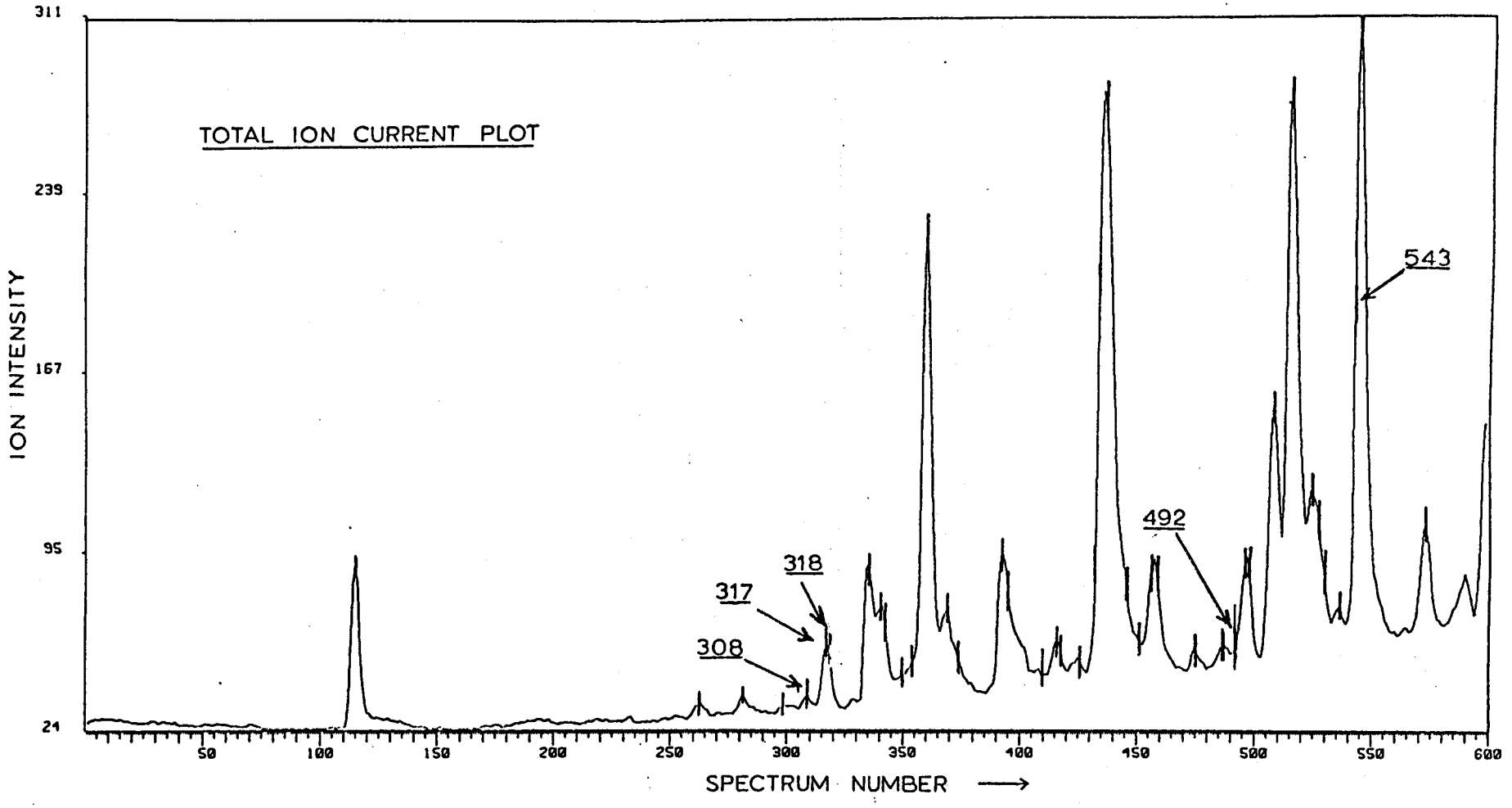
TOTAL ION CURRENT PLOT
311
239
167
95
24
ION INTENSITY
50
100
150
200
250
300
350
400
450
500
550
600
SPECTRUM NUMBER ⟶
308
317
318
492
543

FIGURE 2

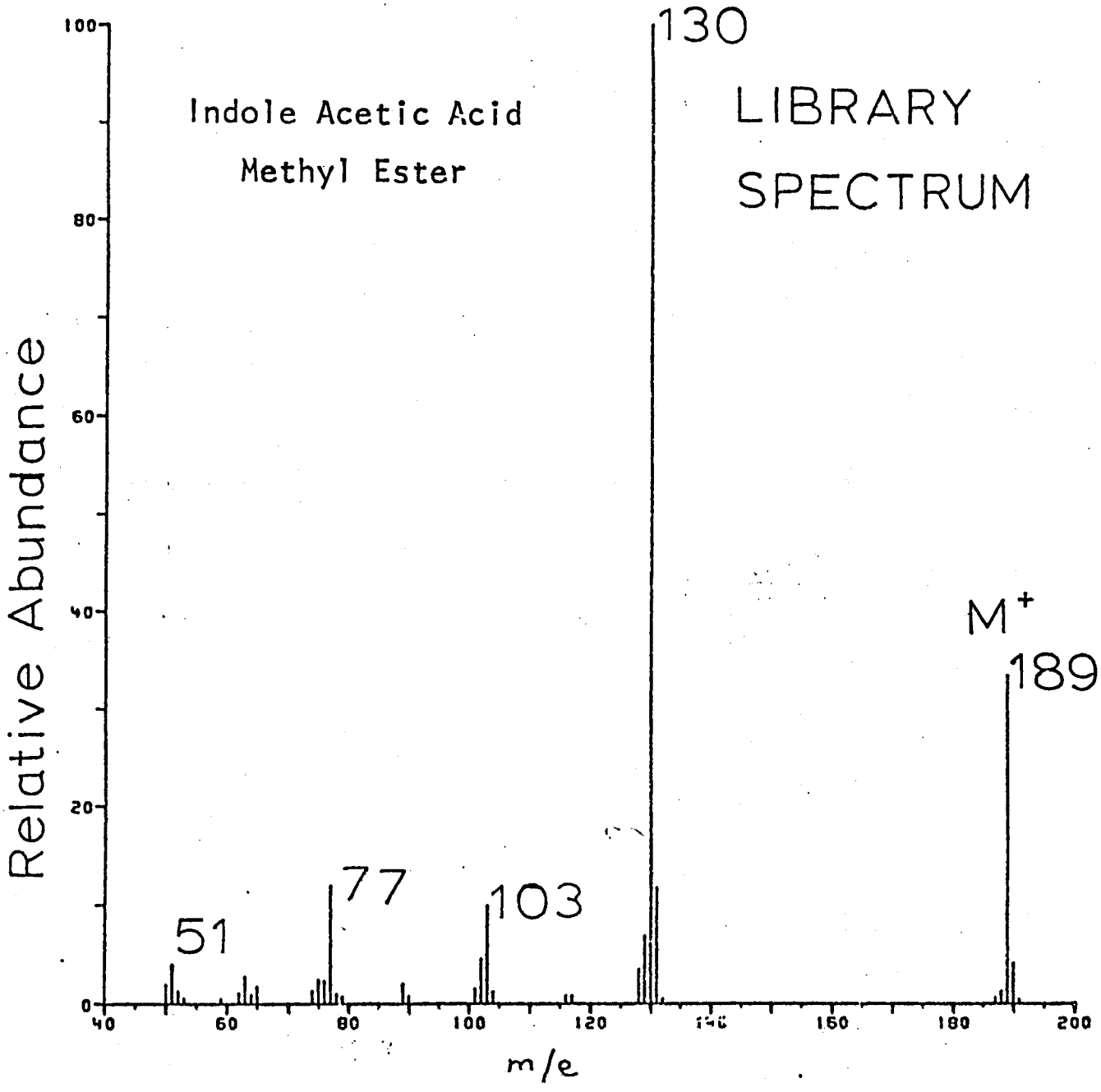
Indole Acetic Acid
Methyl Ester
LIBRARY
SPECTRUM
130
M+
189
77
103
51
Relative Abundance
m/e
100
80
60
40
20
0
40
60
80
100
120
160
180
200

FIGURE 3

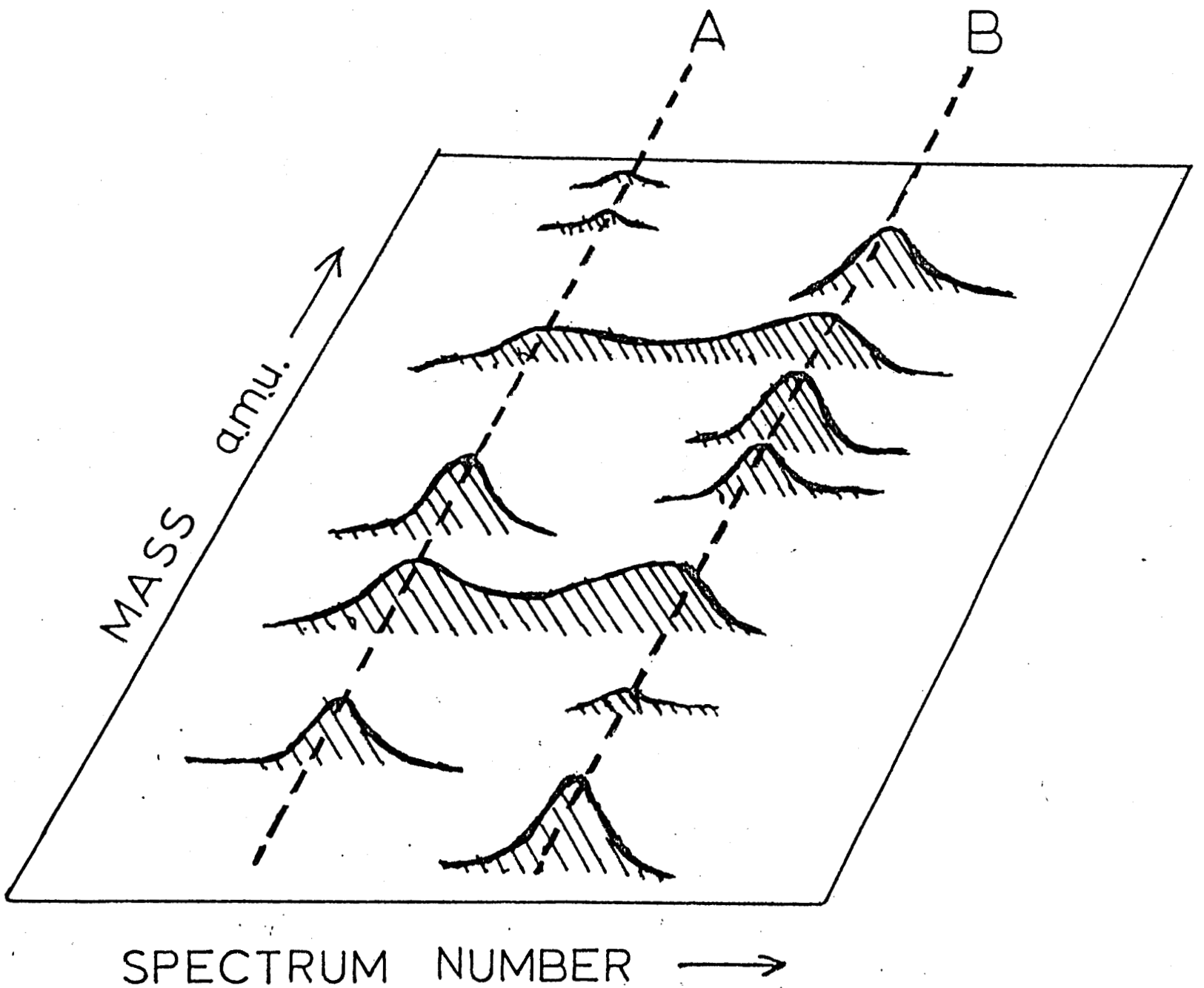

FIGURE 4

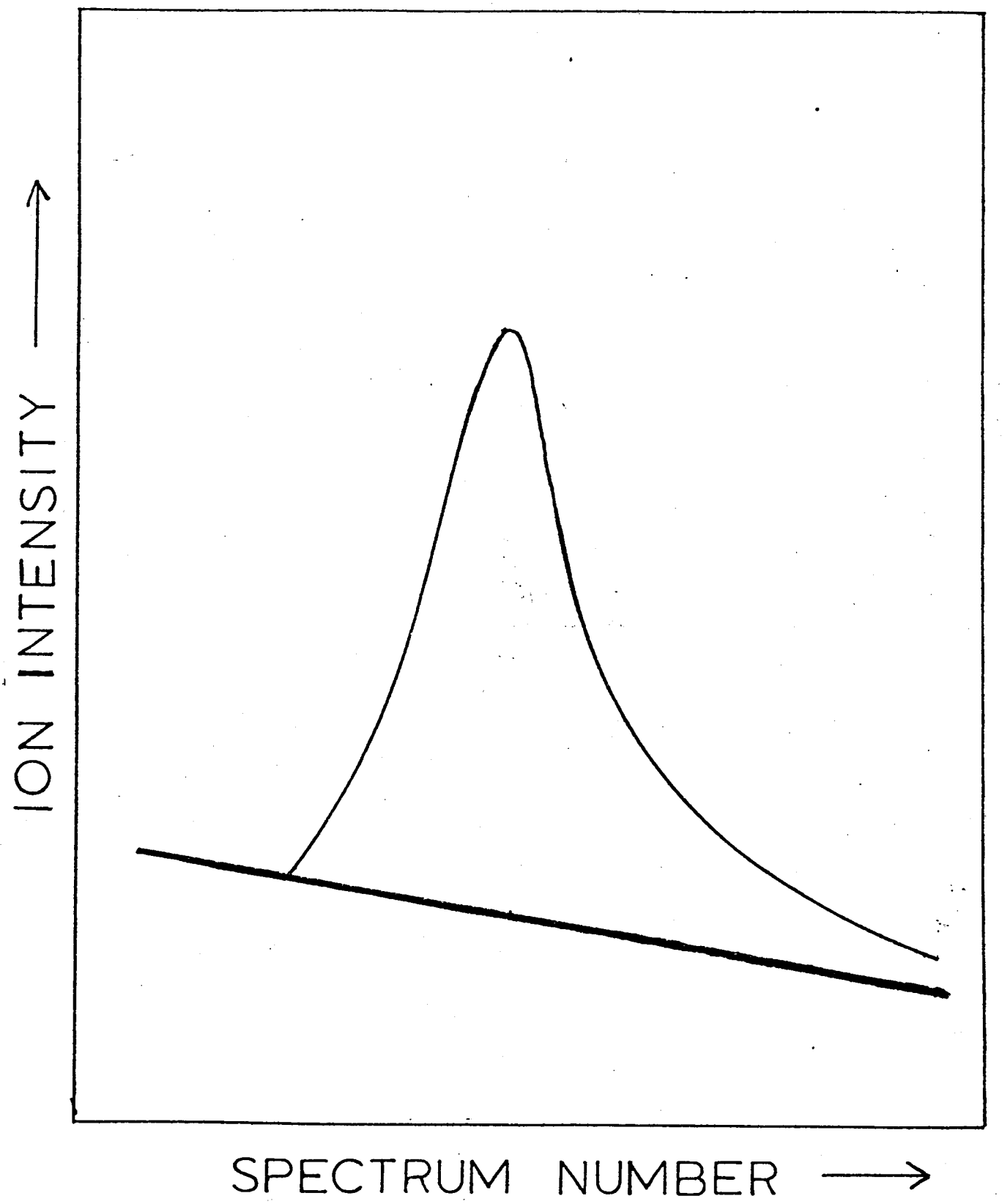

FIGURE 5

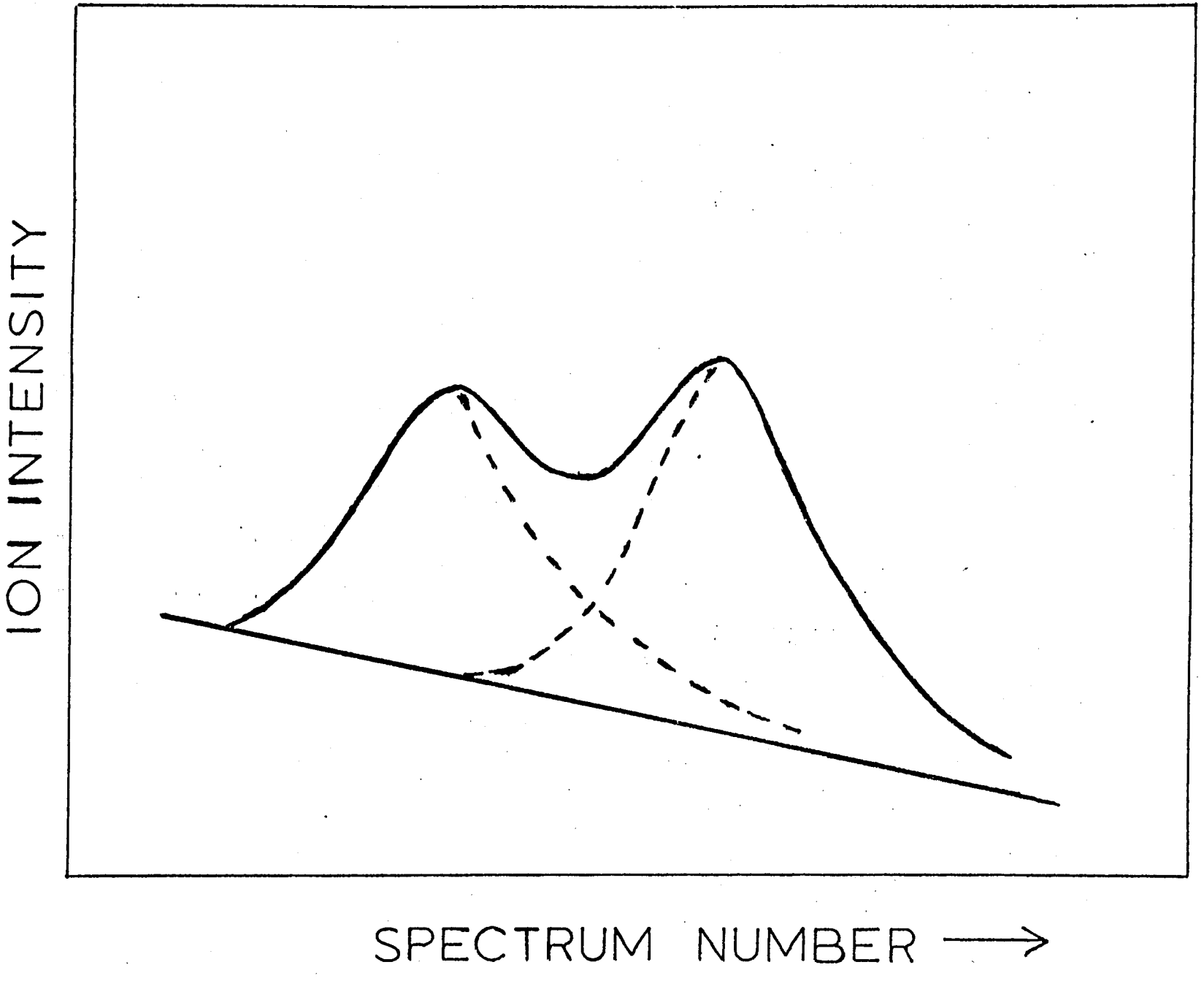

FIGURE 6

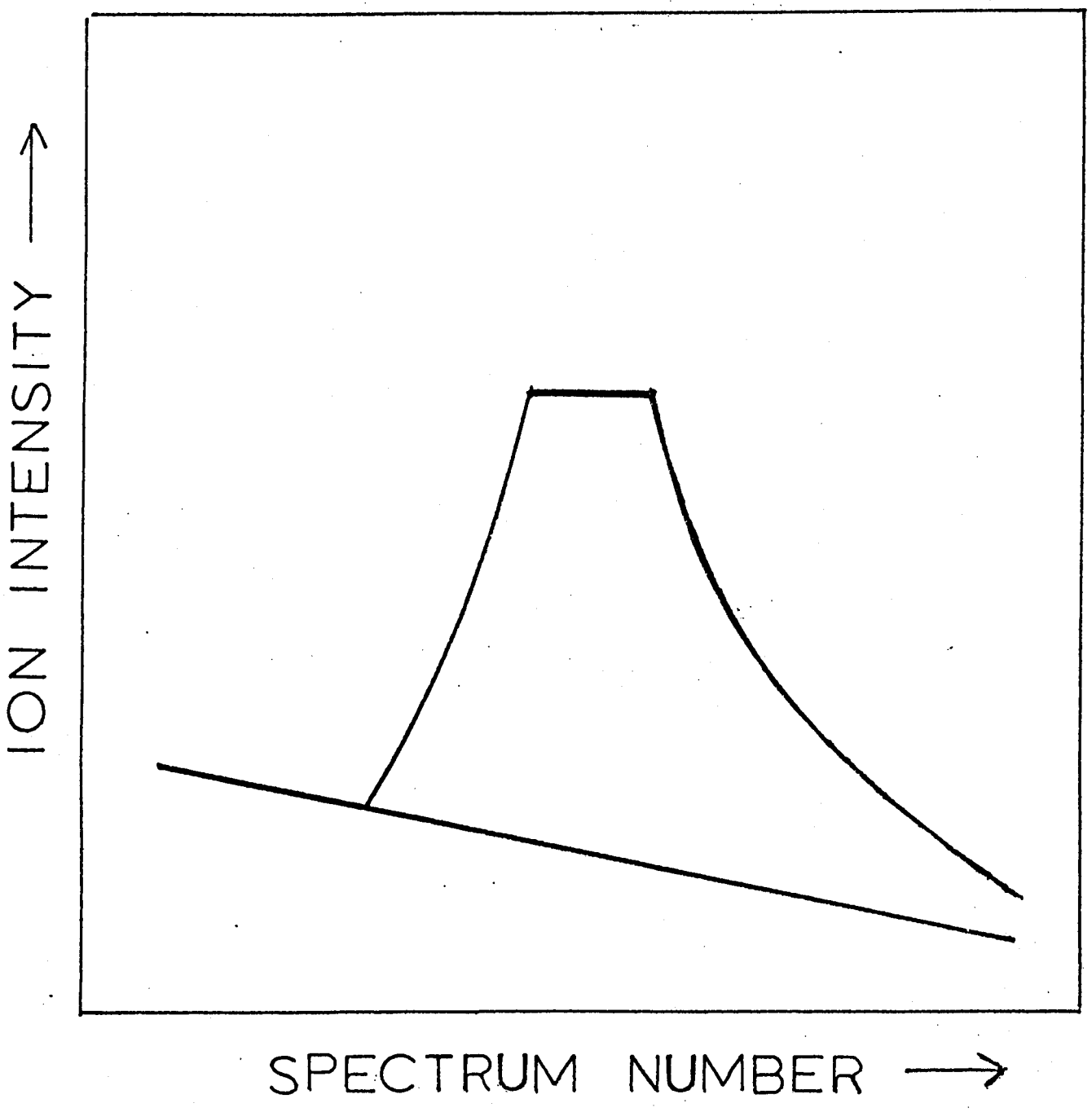

FIGURE 7

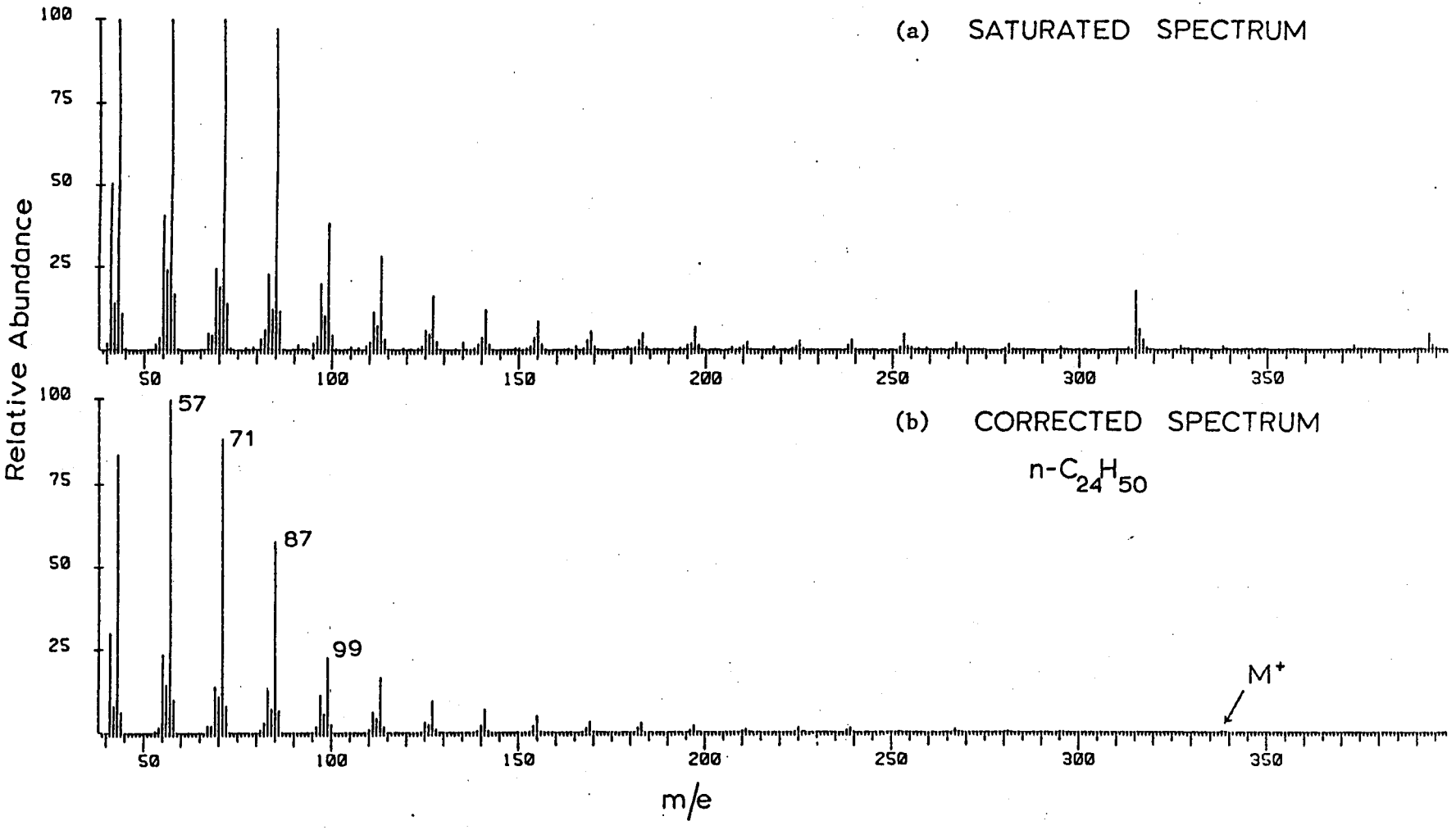
(a) SATURATED SPECTRUM
(b) CORRECTED SPECTRUM
n-C24H50
Relative Abundance
m/e
100
75
50
25
50
100
150
200
250
300
350
57
71
87
99
M+

FIGURE 8

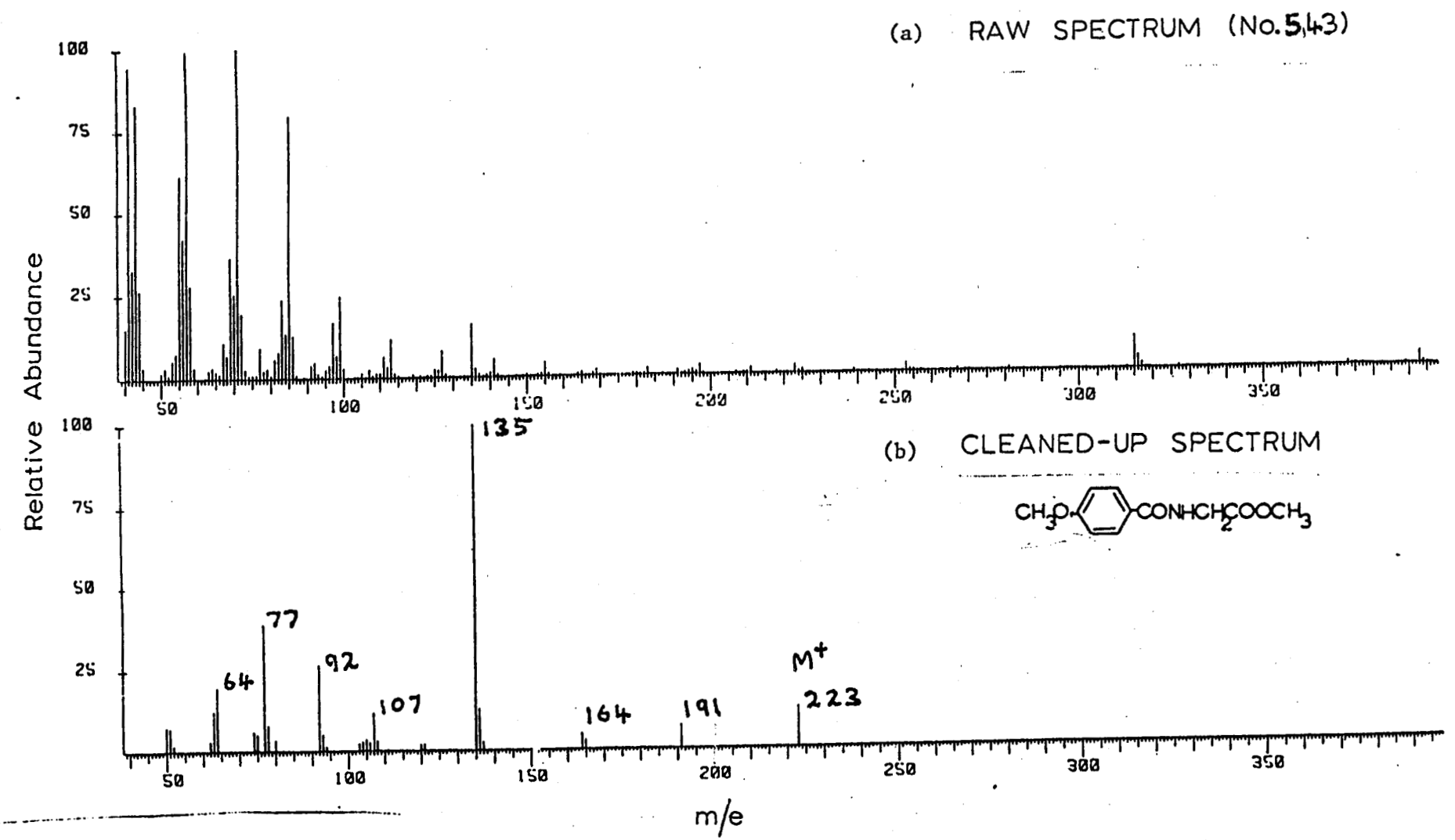
(a) RAW SPECTRUM (No. 5,43)
(b) CLEANED-UP SPECTRUM
Relative Abundance
m/e
100
75
50
25
50
100
150
200
250
300
350
135
77
64
92
107
164
191
M+
223
CH3O
CONHCH2COOCH3

FIGURE 9

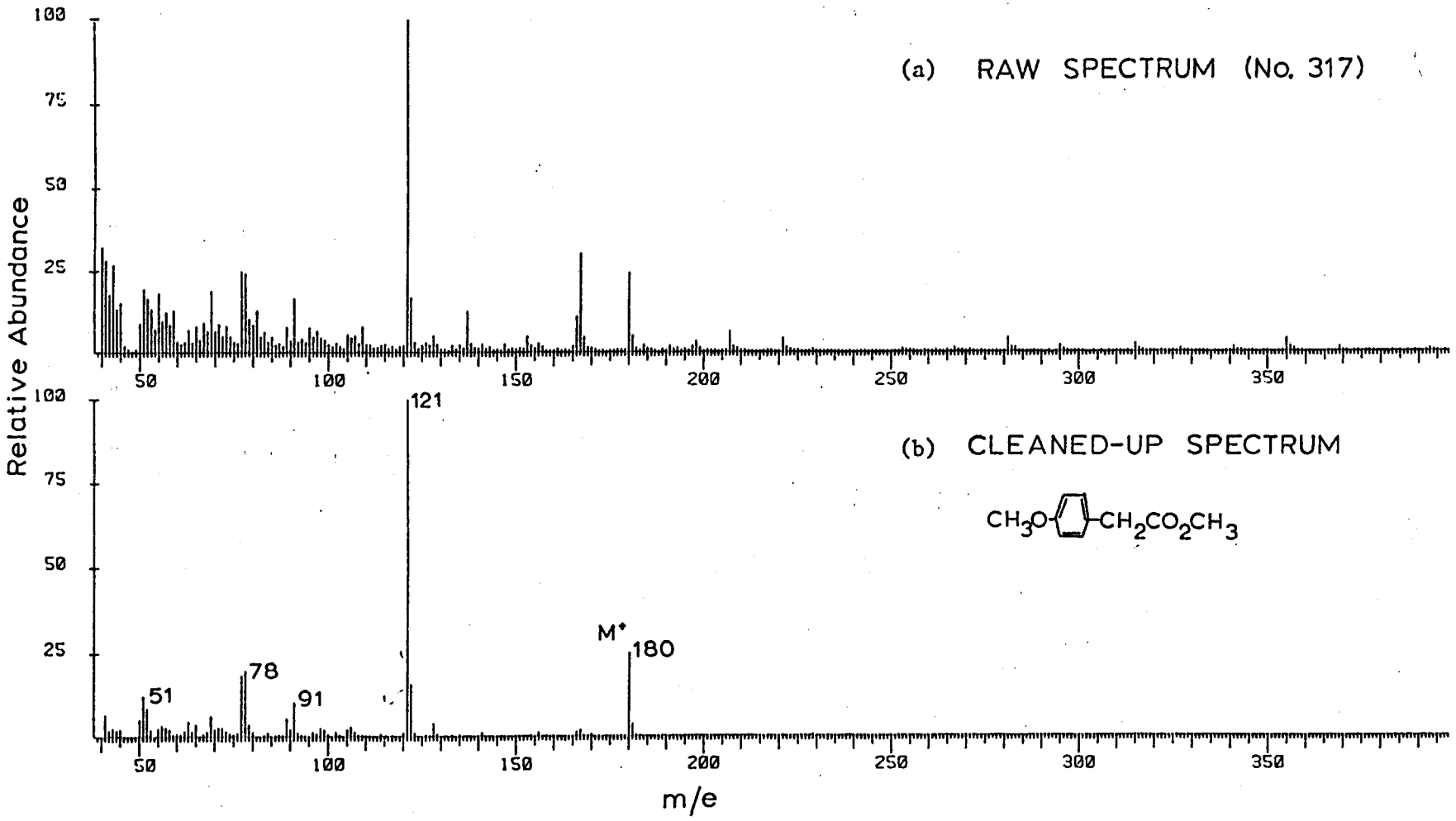
(a) RAW SPECTRUM (No. 317)
(b) CLEANED-UP SPECTRUM
CH3O
CH2CO2CH3
Relative Abundance
m/e
100
75
50
25
50
100
150
200
250
300
350
121
M+
180
78
51
91

FIGURE 10

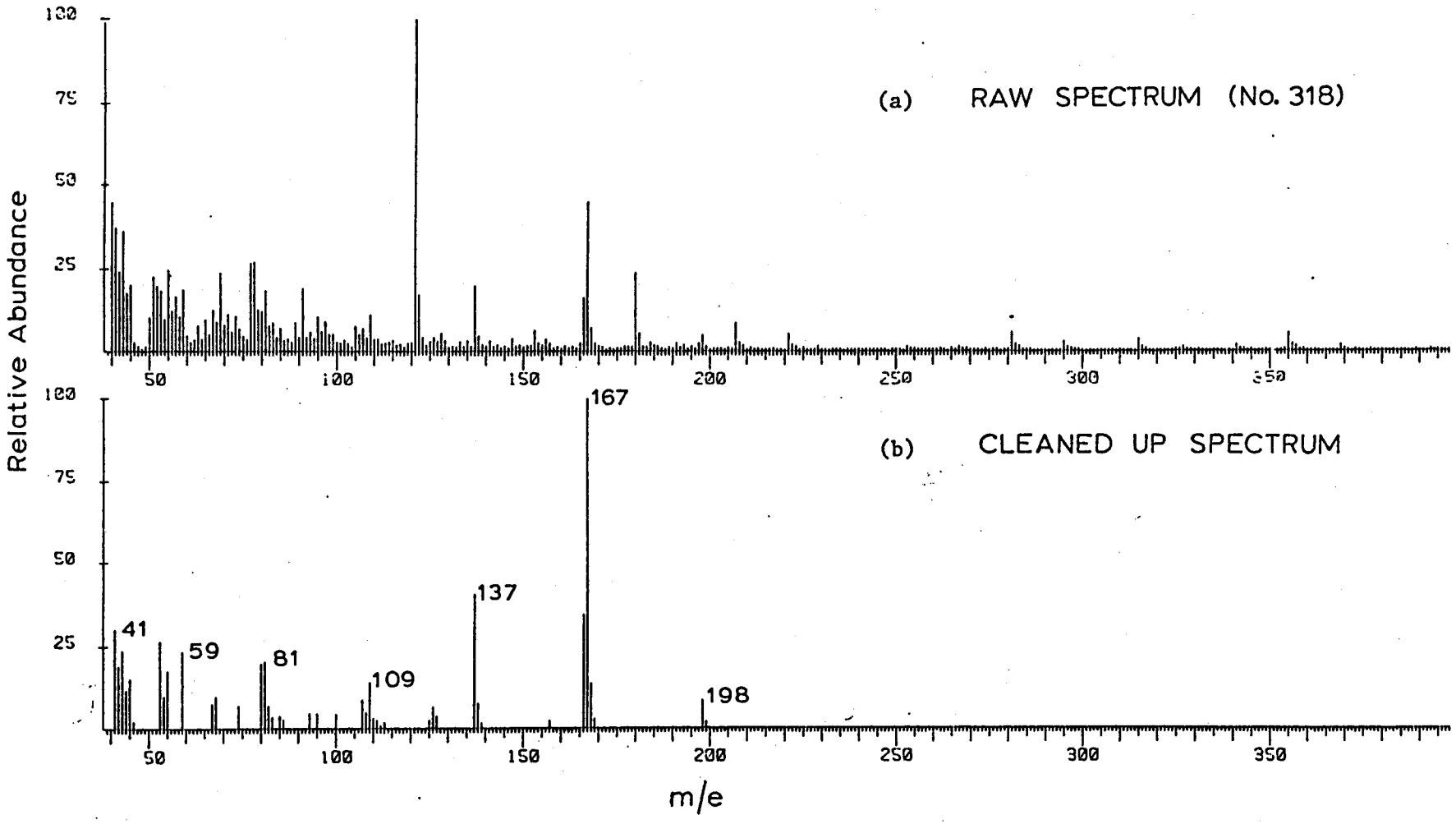
Relative Abundance
100
75
50
25
(a) RAW SPECTRUM (No. 318)
50
100
150
200
250
300
350
(b) CLEANED UP SPECTRUM
167
137
41
59
81
109
198
m/e

FIGURE 11

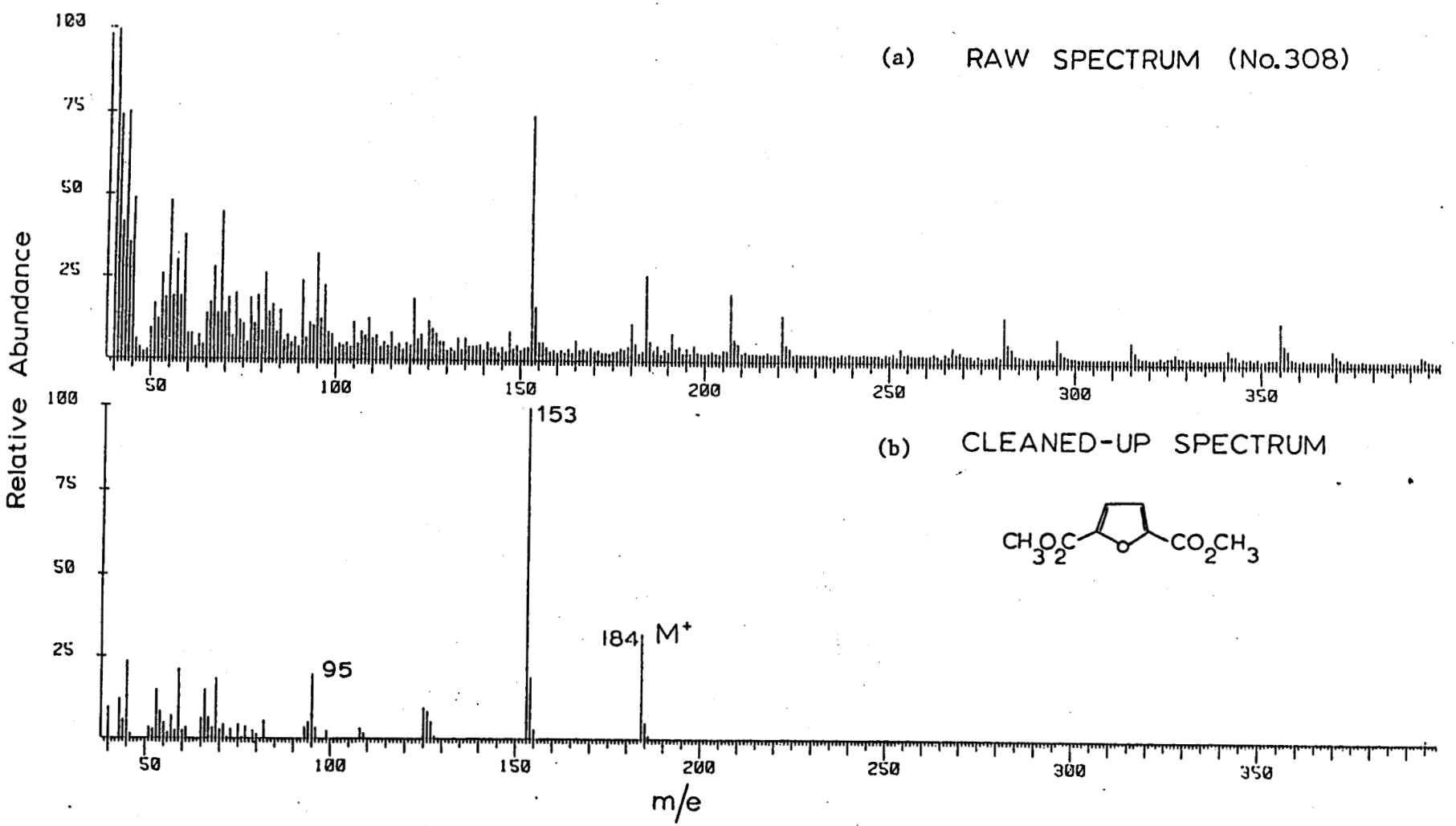
(a) RAW SPECTRUM (No.308)
(b) CLEANED-UP SPECTRUM
Relative Abundance
100
75
50
25
50
100
150
200
250
300
350
153
95
184
M+
m/e
CH3O2C
CO2CH3
O

FIGURE 12